TAMPERE UNIVERSITY OF TECHNOLOGY

ABU FAZAL MD SHUMON

TOOLS FOR SENSITIVITY ANALYSIS IN BIOPROCESSING

Master of Science Thesis

Examiners: D.Tech. Tommi Aho and
Prof. Kari Systä
Examiners and topic approved in the
Faculty of Computer and Electrical
Engineering council meeting on
October 3, 2012.

FORWARD

This thesis was done in the Computational Systems Biology (CSB) group, Department of Signal Processing, Tampere University of Technology, Finland.

ALLHAMDULILLAH, praise be to ALLAH. I am grateful to the most merciful, most just ALLAH for giving me the opportunity and necessary intellect to conduct the thesis. I would like to express my heartiest gratitude to my supervisor D.Tech Tommi Aho for allowing me to carry out my research work with full funding from the department. I am so much thankful to him for his continuous support, guidance over time, patience, and believing on me. I would like to thank Shakira Hassan for her excellent assistance and technical support while setting up the machine for Bioptima planner. I would also like to thank Assoc. Prof. Imed Hammouda for his excellent tips regarding the table of content and writing of the thesis. I would also like to thanks a lot to Prof. Kari Systä for his quick and prompt response with valuable feedback in such a short time in order to proceed for the thesis submission.The master's thesis of M.M. Mahbubul Syeed and Shahed Alam helped me a lot in writing. A special thank for both of them. I would like to express my deepest gratitude towards my parents from the bottom of my heart. They have supported and prayed for me all the way through.

January, 2015, Tampere, Finland

Abu Fazal Md Shumon

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Degree Programme in Information Technology

Faculty of Computing and Electrical Engineering

ABU FAZAL MD SHUMON: Tools for Sensitivity Analysis in Bioprocessing.

Master of Science Thesis, 55 pages
January 2015

Major Study: Software Systems

Keywords: Monte-Carlo Simulation, Sensitivity Analysis, Uncertainty, Tornado Diagram, Violin Plot, Torlin, Decision Analysis.

The domain of bioprocessing and industry of biotechnology is evolving. Remarkable progress in recent time has already brought the industry into a stage of growth. As a result of this, many biotechnological products, foods, and agricultural products are being commercialized. Therefore, there is a transition from research work in the laboratory to the market. In this consequence, bioprocess engineering covers the area from development to the actual production. Different parts of the discipline are being studied and still going on in order to make revenue. Thus, the revenue is concerned, it is important to identify critical phases, uncertain variables involved in those phases, and degree of sensitivity of those uncertain variables. Identifying those variables that plays vital role in production cost makes the decision makers tasks easier. When the key variables have been determined, then the decision makers know where to focus more and analyze further.

In this thesis, we have developed a tool from an existing software platform called Bioptima planner to analyze and identify different sources of uncertainty, uncertain variables, and their level of uncertainty and sensitivity regarding the final produc-tion cost. When significant uncertain variables whose impacts are bigger in the cost have been determined, then the decision makers' tasks become obvious to analyze those variables and leave the less important ones. A case study i.e. Pro-biotic example is used to demonstrate the applicability of the tool. It produces graphical representation of the results using histogram, tornado diagram, and vio-lin plot which serves our purpose of decision making in the context of sensitivity analysis. Significant differences in sensitivities were found among the variables due to different properties of them and the probability distributions applied to them.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Bioprocess engineering (Doran 1995) combines the fields of biotechnology, chemical engineering, and agricultural technology. In contrast, biotechnology processes are used to produce large variety of commercial products such as vaccines, therapeutic proteins etc. (Wang, ym. 2009). Therefore, different branches of the field are eventually being strong sources of making revenue. Especially, the production of recombinant proteins and antibodies has become a major source of revenue during the past few years. Hence, more research has been conducted to optimize the manufacturing cost, identifying critical sources, consistent product, and decision analysis.

Complete living cells or their relevant components those are used to bring desired products or physical or chemical changes can be classified as specific processes i.e. bioprocess (Zaid, ym. 1999). In other words, a process that is able to create an optimal growth environment using the cell culture in a bioreactor (Alford 2006) is identified as bioprocess. Thus, bioprocess operations make use of living cells to produce new products while destroying baleful wastes. Microorganisms are used to convert biological materials for production. Since then, bioprocess is being developed as a lucrative source for an enormous range of commercial products varying from organic solvents to antibiotics. Yeast, enzymes that are industrially useful are also considered commercial products of bioprocessing.

Bioprocessing consists of several processes and phases (Doran 1995) in order to make the raw material to final product. Product recovery, also known as downstream processing is an important part among those processes. This recovery phase is often difficult and expensive due to its sensitive and sophisticated equipment; for some product purification costs 80-85% of the total manufacturing cost. Here, a lot depends on product nature and the broth in selecting actual procedures for downstream processing.

The selection of appropriate operating conditions for bioprocessing is complex due to the large number of interacting stages and variables (King;Titchener-Hooker ja Zhou 2007). Nevertheless, tools to accommodate bioprocess design (SHIOYA;SHIMIZU ja YOSHIDA 1999) and identifying variables sensitivities are of significant. In this thesis, we analyzed and identified different sources of uncertainty and their level of sensitivity. Bioptima Planner software was used as a platform of the analysis work because it contains the functionality needed in process data management and process design. When key variables have been determined, the decision makers may focus on the most significant subset. Probiotic example is considered as a case study. The case study comprise of four unit processes. Inside the processes six materials, four equipment (mainly their usage time), and three product yields were used altogether that are called variables. Extension of the tool varies all those variables within a given range of distributions in order to identify most sensitive variables with their impact.

The structure of the thesis goes as follows. Chapter 1 gives an introduction. Chapter 2 discusses the background work of the thesis. The background consists of application field of bioprocessing, challenges of uncertain parameters, deterministic and Monte-Carlo simulation. Chapter 3 illustrates the approaches of sensitivity analysis in bioprocessing and details description of Bioptima planner. Monte-Carlo simulation extension and its distribution plotting are the main content of chapter 4. Chapter 5 shows how the case study example is varied. Details discussion of the outcome and analogy are described in chapter 6. This thesis concludes in chapter 7 with some conclusive remarks and limitations noted.

2. THEORETICAL BACKGROUND

This chapter discusses about theoretical background of the work. Necessary background and information, their field of application, challenges of uncertain parameters would help the reader to understand and get familiar with the concepts of this thesis.

2.1 Application Fields

Field of biotechnology (Khan 2011) more specifically bioprocess engineering (Mukhopadhyay 2004, Burgess;Osinga ja Wijffels 1999) describes the manufacturing processes using materials, product yields, and other related components with different processes such as centrifugation, chromatography, cooling, formulation, seed fermentation, product fermentation etc.

Biotechnology, chemical engineering, and agricultural biotechnology (Khan 2011) combine a specialized field i.e. Bioprocess engineering. This field is responsible for the design and development of equipment such as Fermenter, Lyophilizator, and Grinder etc. and processes for product manufacturing such as food, probiotics, and polymers. Furthermore, it combines mathematics and biology along with industrial design depending on the purpose. Bioprocess engineering requires study of different processes and techniques those are in use in producing large scale biological product for industries. These techniques help to optimize yield and quality of final product. In broader domain, bioprocess engineering may include principles of mechanical and electrical engineering.

Bioprocess operations transform microbial and plant cell components to manufacture products and demolish wastes. Considering this production scenario by using biological materials, bioprocess has been developed for a large domain of commercial products. This comprises from relatively cheap materials for example organic solvents to costly chemicals as probiotics, and antibiotics etc. Enzymes, living cells, yeasts are considered as commercial products of bioprocessing. Yeast has a significant contribute in developing biotechnological processes and recently yeasts are making extraordinary advances in the knowledge of bioprocessing systems and production levels (Mukhopadhyay 2004). As new areas are emerging, the knowledge is these areas are being expanded really quickly which offers to develop novel products or improvements of methods. However biological systems and structures of internal might be complex and difficult to handle; nevertheless, they follow rules of chemistry and physics and therefore cooperative to engineering analysis. All sort of engineering angles such as design and operation of bioreactor, equipment of product recovery, automation process, production and deployment of products are required in bioprocessing.

It is hard and complicated to select appropriate operational conditions for simulation model including bioprocessing because of large number of cooperating stages and variables. One process in the model would for it is previous process to finish in order to further. Each processing stage usually deals significant amount of variables thus the overall model consists of a noticeable amount of variables. Due to the typical interactions between unit processes and variable’s value (e.g. product yield), it affects the consequent processes. Thus it is important to think model as a whole and identify the sensitivities of the variables for the optimal output. Furthermore, demonstration of performance needs to be satisfactory in likely operating range. Therefore tools to analyze the sensitivities and uncertainties of the variables are vital which will assist bioprocess design.

Sensitivity analysis (SA) is closely related to the bioprocessing engineering. Sensitivity analysis techniques express and explore how model output is dependent upon all of the input variables. SA determines the relative effects and brings up the prominent sensitive variables. Global (King;Titchener-Hooker ja Zhou 2007) or local sensitivity analysis is depending on the model necessity.

If the strength of the interactions between of variables of the model needs to be determined, then global sensitivity analysis techniques could be useful.

2.2 Production procedures in biotechnological processes

Process sequence i.e. a sequence of unit processes in a production scenario provides necessary information related to convert raw materials to final product. A scenario contains list of unit processes. Each of these unit processes contains specific information about equipment, usage time, working time, initial amount, final amount etc.

Downstream processing refers to recovery and purification of biological product. Typically downstream processing deals with sequences of unit processes. This processing in bioproduction consists of several steps, however in general it can be categorized in sections as: preparation, capture, purification, and polishing.

Preparation means decomposition in this context. For example separation of cells from fermentation broth would pass through filtration, centrifugation, microfiltration, and ultrafiltration. Homogenization or leaching kind of process required recovering products from solid sources (plant and animal tissues) can be added in the preparation phase.

Because of many ingredients may lose their biological impact, it is important to get into the capture stage quickly. Water plays vital role of chief impurity here for most of the products. Therefore capture stage removes most of it. Volume reduction is handled, and focus is on the product. Volume reduction is essential for reducing equipment size in further steps. Aqueous two-phases, solvent extraction, adsorption, extractive distillation are some of the unit processes related in this stage.

Contaminants that resemble the product very closely in physical and chemical properties are removed and concentration is increased in purification stage. This stage is expensive due to it is required sensitive and sophisticated equipment. A significant impact in overall downstream process expenditure is because of this stage. Some example process of this stage includes evaporation, reverse osmosis, fractional precipitation, crystallization etc.

Polishing is the final processing stage. This stage lead to the final product with packing of the product that is stable and easily transportable. Typical unit processes here are desiccation, spray drying, drum drying, tray drying, and lyophilization. Some other process such as depyrogenation may be included here that would comprise product safety, depending the use of the product.

2.3 Challenge of uncertain parameters

Usage of few parameters cannot be determined unless we implement that process. In real time operations of bioprocessing, uncertain parameters are being operated. Implementation of a process allow us to know for example the accurate running time, exact number of batches used, amount of material used and so on.

2.3.1 Common sources of uncertainty

Uncertainty analysis detects the uncertain variables those are involved in decision making problems where observations and models represent the context. It aims to contribute technically to decision making aspects through those relevant uncertain variables. In real-time experiments uncertainty analysis deals with measurement. In case of measurement, designed experiment to determine an effect may vary due to instrumentation, methodology and so forth.

Few parameters may not represent real circumstance properly; as we know reality is much more complex, but at-least few predictions can be taken into account to deal with while making decisions. In our case there are different sort of uncertain parameter would appear, where some common sources of uncertainties are described below.

2.3.1.1 Material prices

Material price is the prices of materials being used in unit processes. Prices can be for example in €/g, €/L, €/m2 or €/pc. Material price is used to calculate material and utility cost which eventually added up to the final pricing. Material price is an uncertain parameter, as change of price affects the final production cost.

2.3.1.2 Product yield in fermentation

Product yield in fermentation process is the amount of product being produced by this process, and depends on the quantities of the reactants used in this process. The maximum amount of product which is supposed to be produced in an efficient process could be defined theoretically. But as we see, such perfect process does not exist in reality. Thus, actual product yield towards the end of the process is usually less than what is being calculated in theory. However, factors in the fermentation such as bacteria which are producing the product, make the product yield itself uncertain. Bacteria are sensitive in terms of substance, temperature, and pH in their cultivation conditions. Due to their sensitivity, expected amount of product yield might not be achieved which would eventually effect the final production cost. This is the reason product yield in fermentation is considered as a source of uncertainty.

2.3.1.3 Product recovery rates in downstream processing

In Downstream processing (DSP), product needs to be recovered in order to purify and polish, after passing through several phases. Recovery and purification (Mukhopadhyay 2004) consists of several stages like cell removal, evaporation, extraction, drying etc. Product recovery is crucial in contrast to separate those contaminants that resemble the product very closely in physical and chemical properties. These different stages make the recovery sensitive because of the chemical and mechanical processes the product passes through. The percentage of impurity in the processed materials varies from product to product. All are not in the same scale. This reason causes the variety and uncertainty which consequently varies the final production cost. This is why product recovery rate is one of the sources of uncertainty as well.

2.3.1.4 Operating time

Time of operation or the operating time required to complete a full cycle of the production in downstream processing stages is also source of uncertainty. In previous section (i.e. 2.3.1.3) we have already discussed that each production consist of several stages and each stage has own operation time to come up with the output. Operation times are estimated beforehand. But sometimes it needs to be adjusted due to different variations in the DSP. Suppose offset of impurity of a material is high, and that needs to be reduced because the purification of that particular material requires certain level to produce the desired output. Thus the processing time increases in specific unit process in which material is being used and eventually it propagates toward the final production. If the operation time varies, then it automatically reflects in the final output.

2.4 Deterministic Simulation

Deterministic simulations do not contain random variables. These kinds of simulations mostly consist of equations; input variables are known and they anticipate result as a set of output. Different models are built to perform different simulations. Simulation is a process which acts or mimics a real world event or situation. It is a well-known methodology to find solution of many real-world problems.

For example, a mathematically deterministic model is a representation of $y = f(x)$ (What is a deterministic model? ei pvm). This equation helps us to predict values of y based on x. This model is identically similar to our model we have implemented to anticipate final production cost. We have considered cost of final production in terms of y, where y depends on the function $f(x)$. Function $f(x)$ consists of summation of the uncertain variables. Consider here that the term "anticipation" does not necessarily occur in the past, future, or even the present. We simply interpret it as "what-if" situation, and the aim is to empirically estimate the sensitivity. It helps us identify what is the sensitivity in the output for a particular x or x's.

User specifies input variables in the deterministic simulation models such as values of the variables, their distribution types, range of the probability distribution, and yields output. Our approach has been to represent information regarding output by Tornado diagram, violin plot, and histogram.

2.5 Monte-Carlo Simulation

Anticipation or prediction is a usual part of every decision making. Uncertainty, ambiguity, and variability are faced literally. However, sometimes future cannot be predicted accurately due to unprecedented access to information. Monte Carlo simulation, also called Monte Carlo Method (Monte Carlo Simulation and Methods Introduction - Goldsim 2013, Monte Carlo mathod - Mathworld 2103, Kalos ja Whitlock 2008) as well lets to preview possible outcomes of the decisions and assess the impact of variability, and consents to make better decision under uncertainty.

Monte Carlo simulation (Rubinstein ja Kroese 2008, Monte Carlo SImulation - MATLAB and SIMULINK 2013) is a technique used in mathematical or statistical problems by using repeated sampling to determine the properties of some phenomenon. This simulation uses heuristics to calculate probable outcomes which are going to happen most likely. Obviously the application of it varies from field to field such as manufacturing, computational biology, applied statistics, artificial intelligence; and has been used to model a variety of physical and conceptual systems. However, optimization, numerical integration, and samples generation from probability distributions are identified as distinct problems where Monte Carlo simulation can be used mainly. In our model this simulation is used in terms of probability distribution to generate samples in certain range. This sort of simulation techniques (Doubilet, ym. 1985) helps understanding the impact of uncertainty and variables that are important component of decision making (Critchfield ja Willard 1986).

Monte Carlo simulation works in a way such as by substituting values of a given domain and anticipating the possible results. Each time it iterates over and takes a set of random values of variables from the probability distribution function. Then it computes the results. Depending on the uncertain variables and their specific ranges, a Monte Carlo simulation would perform enormous amount of calculation before it is done.

Range of values can be estimated in some cases, for example distribution of possible values through mean and standard deviation of returns in a production line. And more realistic image of future can be predicted by using possible range of values. In such case probability distributions (Ang ja Tang 2003) are a much more realistic way of producing these range of values and describing sensitivity and uncertainty in variables. Thus as obvious, result of this kind of range of estimation would also be a range and show the both extreme: high and low along with the middle path. The essence of range estimation can be understood if we consider an example. Suppose total production cost of “Probiotic example” is 173000 euro being calculated based on the default values of the variables. But if the total cost would have been calculated based on the ranges of the values of the variables, instead of default ones, then we would have the range of total production cost including minimum and maximum. In such case optimal cost could be chosen. Following distributions are used in our model to estimate ranges of values:

Normal distribution: mean and standard deviation (variation about the mean) are defined by the user. Most likely occurring values are in the middle or nearly in the middle. This distribution is symmetric and it describes events like height of peoples.

Uniform distribution: both minimum and maximum are defined by the user. In this distribution, all of the values have equal chance to be occurred. Manufacturing or production cost would be an example where variables could be distributed uniformly.

Triangular distribution: minimum, canter (i.e. mid value), and maximum values are defined by the user. Values those are mostly around the center are most likely to happen.

Monte Carlo simulation is considered because of it is simplicity. It is easy to understand which inputs had the biggest effect on bottom-line results.

3. APPROACHES IN SENSITIVITY ANALYSIS IN BIOPROCESSING

This chapter discusses about the basic concept of sensitivity analysis, its methods of different kinds. Each method is described briefly.

3.1 Sensitivity Analysis

In principle, Sensitivity Analysis (SA) (Saltelli, ym. 2008) is a simple straight-forward idea; change the input parameters and observe the effects in the final output. SA determines how sensitive the output of a mathematical or statistical model or simulation is, in response to the changes of the input parameters of those models or simulations which are uncertain (Francois, ym. 2006). However SA can be defined as a technique used to determine how changes in the values of different input variables affect the output under a given set of assumptions. This technique is used within specific range of one or more input variables. For example, changing the price of a material in a unit process (Jochen, ym. 2011) would affect the final cost of production. This impact of changing material price would indicate how important the material was, in terms of uncertainty. SA is possibly one of the most common choices to the modelers who wish to support decision makers. The importance of SA is being acknowledged as:

"A methodology for conducting a [sensitivity] analysis ... is a well-established requirement of any scientific discipline. A sensitivity and stability analysis should be an integral part of any solution methodology. The status of a solution cannot be understood without such information. This has been well recognized since the inception of scientific inquiry and has been explicitly addressed from the beginning of mathematics". (Anthony 1983)

Sensitivity analysis can be executed by modelers for several reasons including: additional research of a parameter, insignificant parameters that can be overlooked, parameters responsible most to output variability, highly output correlated parameters.

“ (Crick;Hill ja Charles 1987) have made a distinction by referring to 'important' parameters as those whose uncertainty contributes substantially to the uncertainty in assessment results, and 'sensitive' parameters as those which have a significant influence on assessment results.”

Sensitivity analysis can serve different useful purposes including robustness of the results of a model, Identifying threshold values of a model, assumptions for the decision makers (David 1997). Along with the above, SA also finds the region of input factors that causes maximum or minimum output of the model. The maximum and minimum is defined with respect to the default value, for every input parameter that is being varied. The next section of this chapter will discuss more details about the maximum and minimum values.

There are different types of methods (Bellman ja Astrom 1970, Iman ja Helton 1991, Fery ja Patil 2002) to analyze sensitivity and uncertainty. The literature contains more details of these types. A comprehensive review has been provided here (Hamby 1994). In this thesis, methods are classified as: one-way method and multi-way method.

Conceptually the simplest method of SA is one-way method. This method examines effect on the model output by varying only one input variable at a time across its entire range, while holding the others fixed. Sometimes this kind of analysis is being referred as local analysis. All input parameters are randomly sampled with their defined probability density function. While varying one parameter at a time, one random sample being picked from its entire range; and all other parameter list remain unchanged (i.e. default value). Consider the example scenario below:

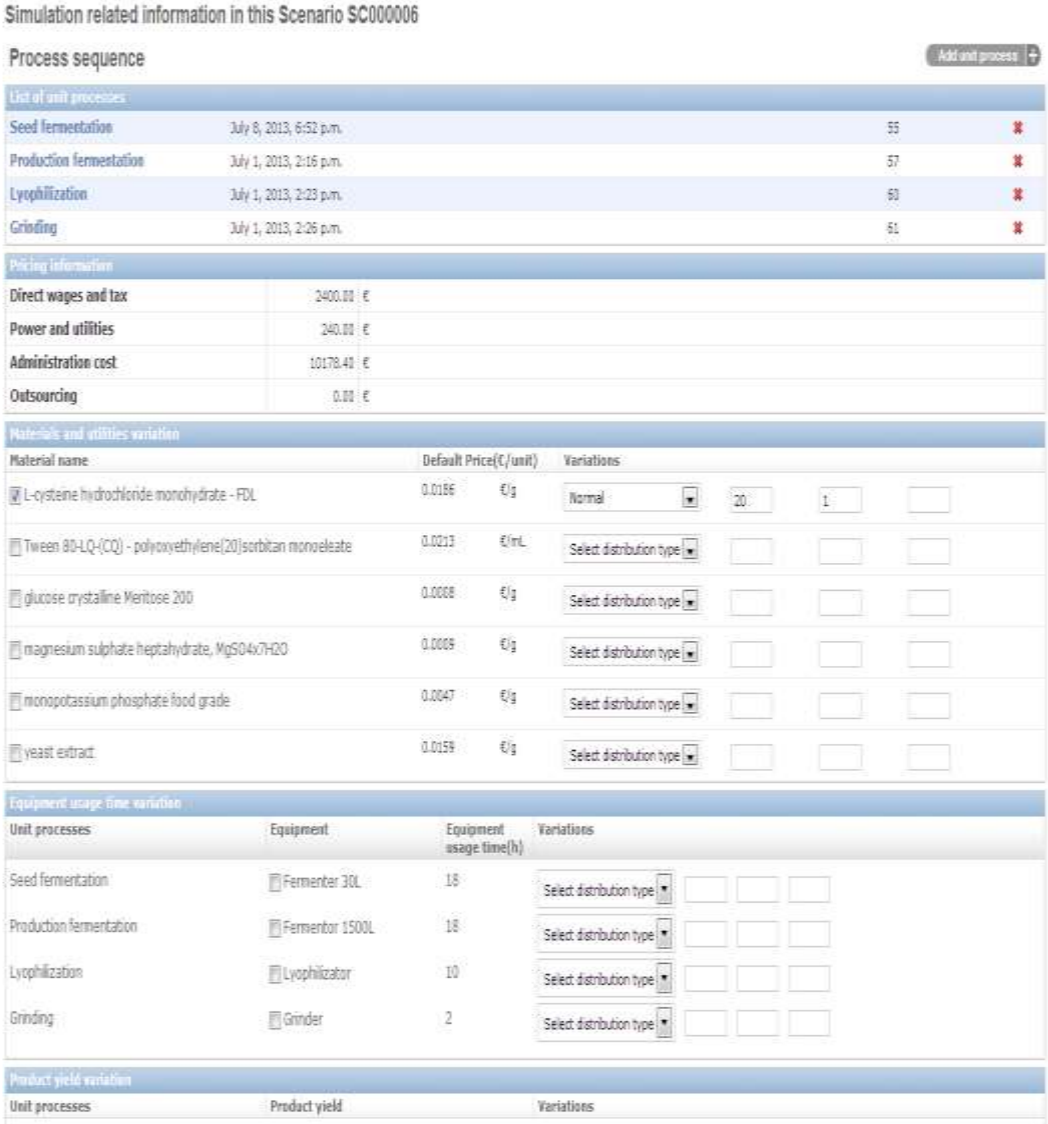


Figure 3.1: Example scenario of one-way method of sensitivity analysis

In this scenario, default price of the material “L-cysteine hydrochloride monohydrate - FDL” are being varied according to normal probability distribution with the mean of 20 and standard deviation of 1. Rest of the prices of the materials and equipment along with their usage time are kept default. The amount of variability introduced into the model output while varying this material’s price would show how sensitive this material is.

Multi-way sensitivity analysis means to evaluate the model output by varying two or more parameters simultaneously. Sometimes examining the relation of changing two or more parameters is necessary. *Figure 3.2* is illustrated an example of multi-way sensitivity analysis.

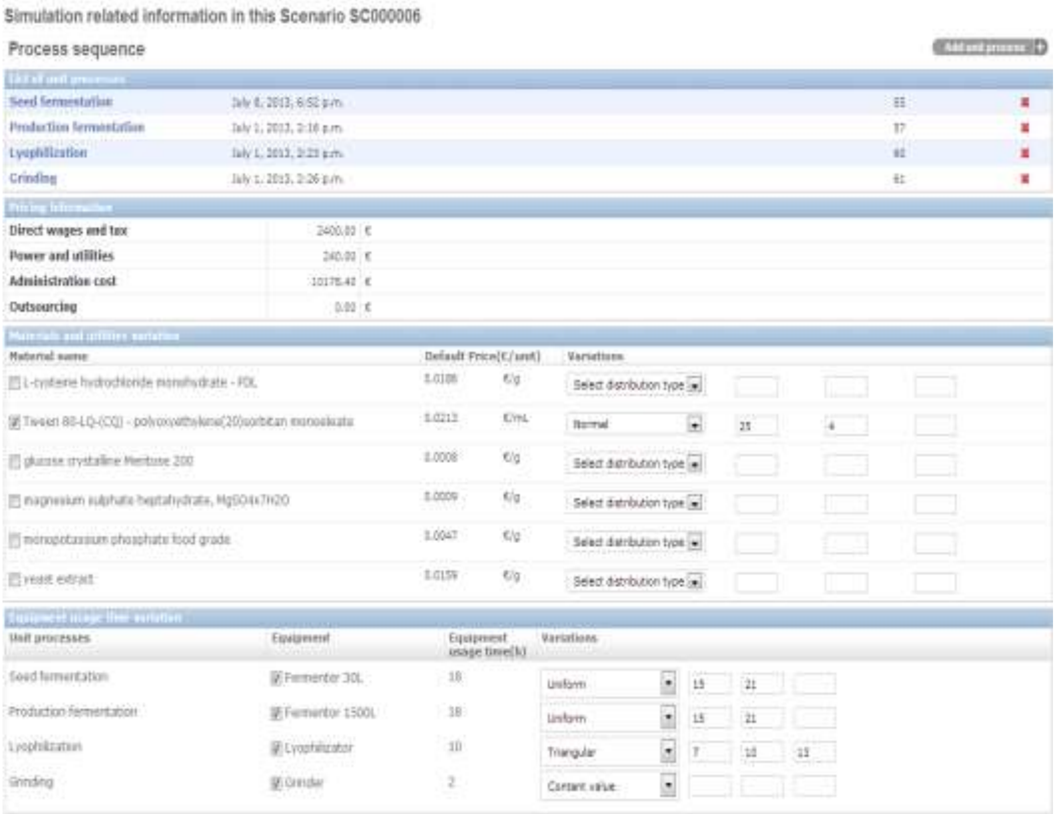


Figure 3.2: Example scenario of multi-way method of sensitivity analysis

Usage time of each equipment (i.e. Fermenter 30L, Fermenter 1500L, Lyophilizator, and Grinder) is being varied according to their probability distribution function (i.e. uniform, uniform, triangular, and constant respectively). The results for each potential combination of values along with other input variables are in the model output.

3.2 Tornado Diagram

Tornado diagram (TD), sometimes called tornado plot or tornado chart or sensitivity char, is a set of one-way sensitivity analysis. It is an established way of showing a one-way sensitivity analysis of several variables in the same output by a TD (Idefeldt ja Danielson 2007). This classical bar chart/graph provides decision makers a quick overview of the analysis involved by plotting the results of various anticipations on the final output. Interpretation of bar depends on the context, for example, in this case bars represent the deviation of the total production cost from the default cost in both side. TD provides the importance of the variables in terms of their sensitivity in the context. A typical TD is shown below:

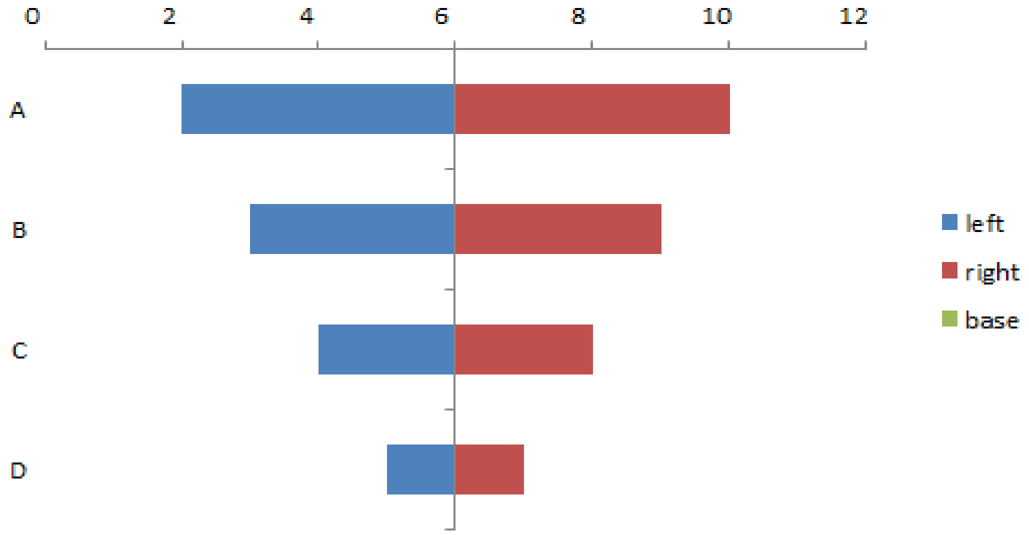


Figure 3.3: Example of a Tornado Diagram.

Vertical line is drawn as base line which means the base value, values in the horizontal line (0…12) are the arbitrary units, and different input variables are labeled as A, B, C, D. The name is appropriate as “Tornado” because it is descriptive, and the final diagram somehow visually resembles one half of or a full tornado. The input variables involved are arranged vertically. The vertical line in the middle of the diagram shows the base value of comparisons. The horizontal bars represent the effects of the variables’ variability to the output. TD organize each bar such that largest bar comes on the top, second largest bar comes second from the top and so on. The upmost variables contain most crucial uncertainties that dominate the output. And decision makers will pay attention to such variables while leaving other related variables as they are.

The advantage of this kind of interval approaches is that the critical variables are already marked and presented visually. For example we consider an example scenario (i.e. Probiotic example) in our Bioptima planner software (See Section 3.3). This scenario consists of 4 unit processes and those processes contain 6 materials all together. Here material price is considered as a common source of uncertainty (Section 2.3.1). Furthermore, equipment usage time and variation of product yield would be considered other sources of uncertainty. So if a decision maker needs to visually compare all those uncertain parameters all together, it would be really tough using normal bar chart. We have considered an example of approximately 14 uncertain variables. But there are other scenarios which may contain much more uncertain input variables. In those circumstances, decision makers task be would nearly impossible. Therefore TD provides great insight here. Moreover important variables appear on the top of the diagram with their impact on the model, which makes decision makers task easy where to put importance to. And lower impact variables eventually come towards the end, i.e. those variables could be overlooked sometimes depending on the context.

3.3 Bioptima planner

Bioptima Planner is software that offers calculation of the production and backend management system. This Planner is accessible through Internet simply using a web browser. The software has centralized data storage, supports multiple users with modifiable user permissions. The users are able to design various manufacturing scenarios and compose offers based on the selected scenarios. The software calculates automatically e.g. the total manufacturing costs and different pricing options. In addition, Bioptima Planner performs sensitivity analyses that show the effect of various variables to the costs.

The following figure roughly depicts the hierarchy of the Bioptima planner:

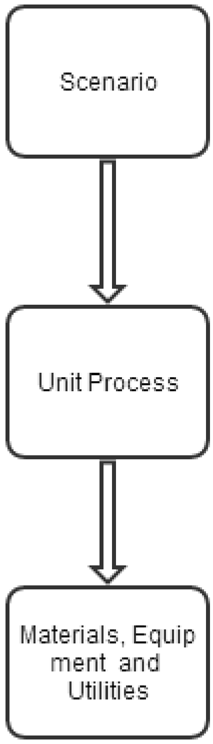


Figure 3.4: Scenario contains unit processes; unit process contains different materials, equipment and utilities

3.3.1 Scenario

A scenario describes the manufacturing process of a production. Any scenario in the planner keeps necessary manufacturing related information (e.g. basic information, list of unit processes, cost summery etc.) and deals how a process sequence (i.e. a sequence of unit processes) can be utilized in order to convert raw materials to end product.

3.3.2 Unit Process

Unit process is a basic step in the process sequence that performs single unit process under some specific circumstances. Unit process involves bringing physical change. This process contains specific information regarding equipment, usage time, initial amount, final amount etc. The user specifies the following required information to run the unit process:

- Unit process name
- Equipment in use
- Working time. It can be specified in hours or “working time share”. Working time share means the proportion of the equipment usage time.
- Outsourcing.
- Some unit processes need additional information which may include Initial volume of medium (L), Product yield (%), Waste volume (L), etc.

3.3.3 Material and Utility

Materials and utilities are specified based on their consumption rates. And the consumption rates are calculated as follows:

- Usage rate. A number and a unit for describing the rate of consumption. The available units are derived from the unit that is given in the list of materials and utilities. For example, if a unit in the list of materials and utilities is “kg” for a particular material then the available units are “g” and “kg”.
- Usage basis
- Recycling / replacement rate and Recycling / replacement dimension. These fields are used to specify the possible recycling or replacement.

3.3.4 Pricing

When a new unit process is created in a scenario, the current price information is retrieved from this list of materials and utilities. The prices in the unit process are updated when the user either edits the unit process and saves it or performs an update by pushing the update button in the scenario summary page. The latter updates the price information in all unit processes of the scenario

3.3.5 Platform and Language

The Bioptima planner is developed using the programming language Python and a high-level python web framework Django. SQL is used for database query.

4. AN INTEGRATED SENSITIVITY ANALYSIS TOOL

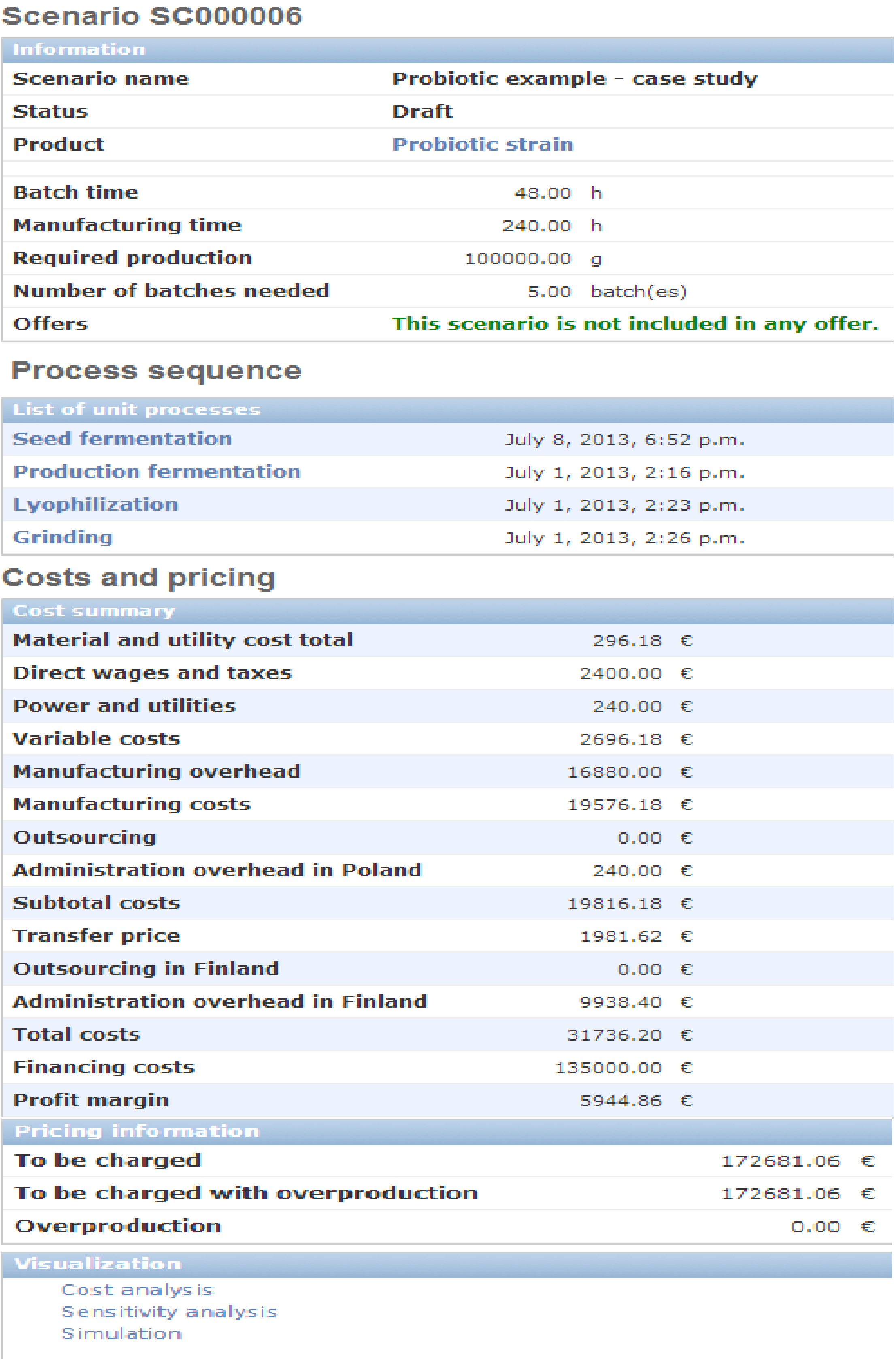
This chapter elaborates the Monte Carlo Simulator extension along with Torlin, and plotting distributions.

4.1 Monte-Carlo Simulator

When any model is designed to predict ahead of time, then there is certain assumptions are taken into account. In our context these assumptions are about the cost of the production. The best it can provide is the most likely values as these are the projections in future. This estimate is drawn based on the distributions applied onto the variables. Estimate is obviously useful, but it brings some intrinsic uncertainty. Therefore, most uncertain variables and their sensitivity impact are identified.

This is how production cost of probiotic example appears (figure 4.1) with default values of the variables. This provides us some room to estimate a range of values in some cases. Based on the information available and the uncertain variables we can estimate the maximum cost it might take in the worst possible scenario, and the minimum cost in the best possible case. Range of possible values can depict a more realistic picture of most likely happening.

We have already seen the idea of Monte-Carlo simulation briefly in section 2.5. When the simulation link is clicked on the bottom of the figure 4.1, it takes us to the simulator shown in figure 4.2. This simulator is used in our simulation model where we can select the type of the distribution for each corresponding variables. Depending on the type of the distribution values such as mean, standard deviation, and minimum, maximum are being put to the next empty text fields. Once all values are set to be distributed, run button (shown on the bottom of the figure 4.2) is clicked to proceed to the resulting page. This page brings the output of the simulation in terms of histogram and violin plot and generates the data related to tornado diagram. These data are passed into the middleware to produce the desired histogram. A limitation regarding the graphical representation of tornado diagram in the same go with histogram and violin plot is described in section 6.4.



Scenario SC000006

Information	
Scenario name	Probiotic example - case study
Status	Draft
Product	Probiotic strain
Batch time	48.00 h
Manufacturing time	240.00 h
Required production	100000.00 g
Number of batches needed	5.00 batch(es)
Offers	This scenario is not included in any offer.

Process sequence

List of unit processes	
Seed fermentation	July 8, 2013, 6:52 p.m.
Production fermentation	July 1, 2013, 2:16 p.m.
Lyophilization	July 1, 2013, 2:23 p.m.
Grinding	July 1, 2013, 2:26 p.m.

Costs and pricing

Cost summary		
Material and utility cost total	296.18	€
Direct wages and taxes	2400.00	€
Power and utilities	240.00	€
Variable costs	2696.18	€
Manufacturing overhead	16880.00	€
Manufacturing costs	19576.18	€
Outsourcing	0.00	€
Administration overhead in Poland	240.00	€
Subtotal costs	19816.18	€
Transfer price	1981.62	€
Outsourcing in Finland	0.00	€
Administration overhead in Finland	9938.40	€
Total costs	31736.20	€
Financing costs	135000.00	€
Profit margin	5944.86	€

Pricing information		
To be charged	172681.06	€
To be charged with overproduction	172681.06	€
Overproduction	0.00	€

Visualization

Cost analysis
Sensitivity analysis
Simulation

Figure 4.1: Pricing info of Probiotic Example with default values

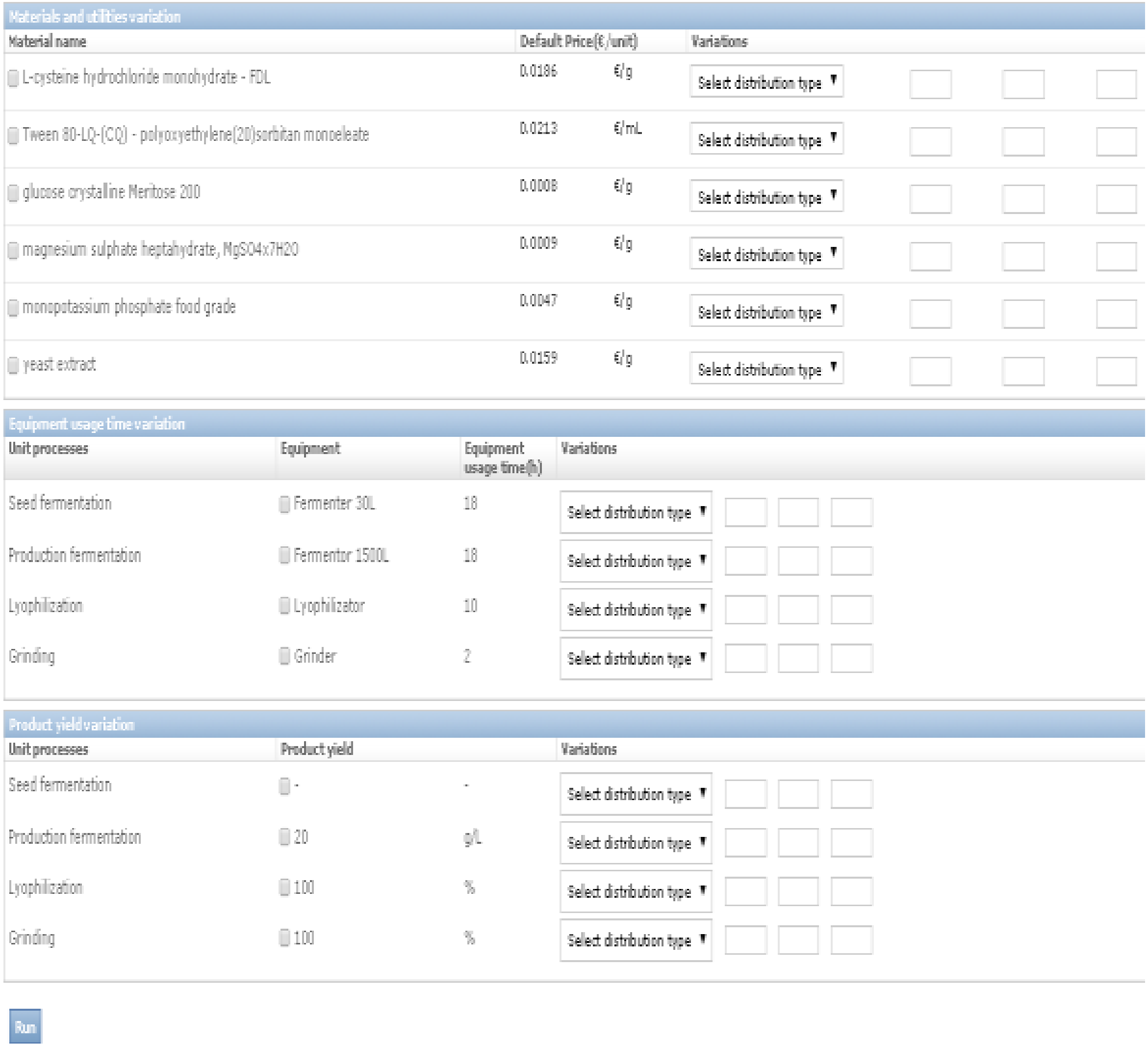


Figure 4.2: Monte-Carlo simulator

4.2 Torlin

Torlin consist of tornado diagram combined with violin plot. We have already seen TD and applications of it in the section 3.2. TD provides standalone impact of a single variable responsible for uncertainty. But violin plot gives an overall exposer of the uncertain parameters.

Numerical data plotting can be done using several methods. Violin Plot (VP) (Hintze and Nelson 1998) is one of them. When differentiation point is concerned, then most likely VP is used instead for different distributions, especially for uniform distributions. VP comprise of plot of kernel density and box plot. More specifically it begins with a box plot and eventually adds up the plot of the kernel density which is rotating to the each side of the box plot. In fact VP is a representation of box plot surrounded by the applied distributions of the data. It helps combining available information to produce the summary in the form of box plot. Apart from similarities with box plot, VP has some uniqueness in itself. It portrays density of probability of data at different levels. Torlin has several violin plots (three plots in our case) in one figure and that gives a summary figure when several variables are varying at the same time. Therefore, this hybrid representation pulls up a quick meaningful comparisons of distributions applied in it. Consider the following example below to have a better understanding of VP.

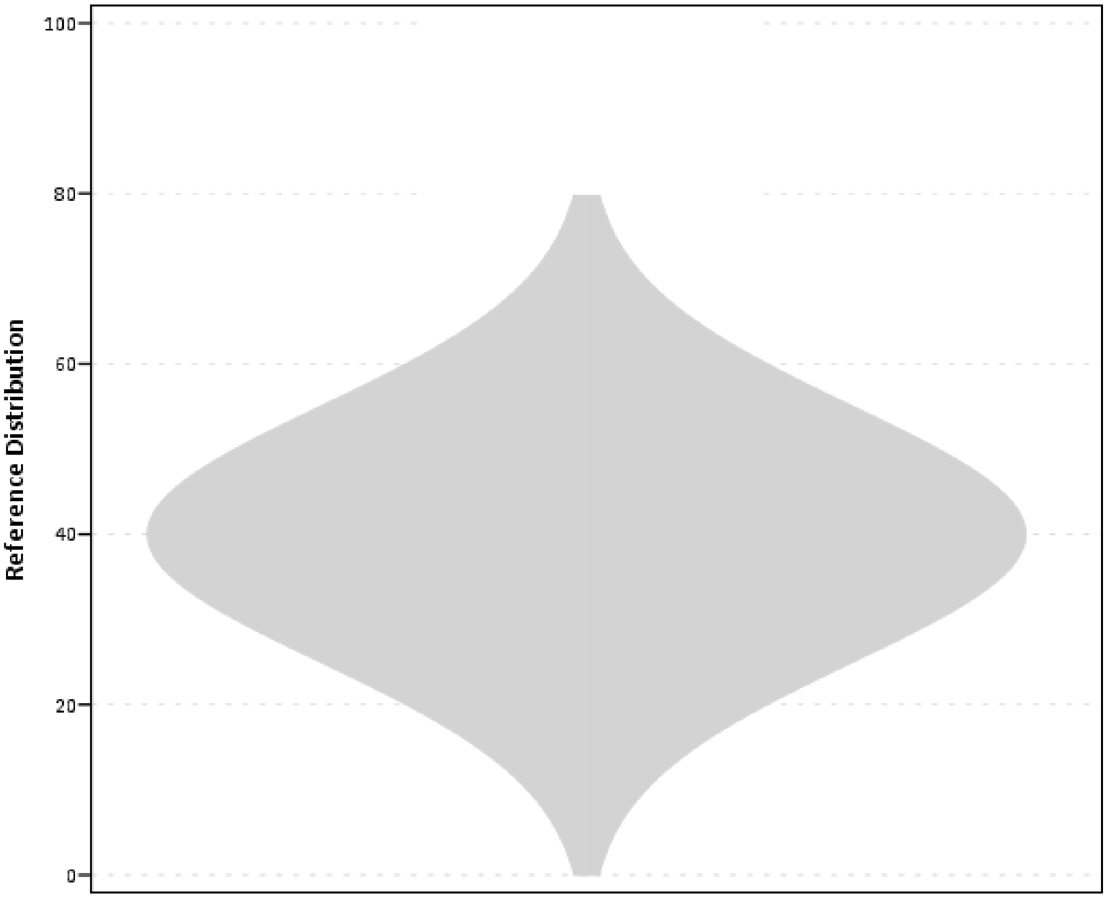


Figure 4.3: Example of Violin Plot.

It is an arbitrary example of VP. Medians are in the center of the plot whereas the top of the plot shows the maximum value and bottom shows the minimum. Advantage of VP over box plot is that it shows density. And when the density is higher, then the wider the VP becomes.

4.3 Plotting the distributions

As it has been mentioned in the earlier (section 2.5) that normal, uniform and triangular distributions are used mainly in this thesis context. Different combinations of distributions were applied such as normal distribution in the material price, uniform and triangular distributions in the equipment usage time, and normal and uniform distributions in the product yield variations. Following are the example of distributions as applied specifically:

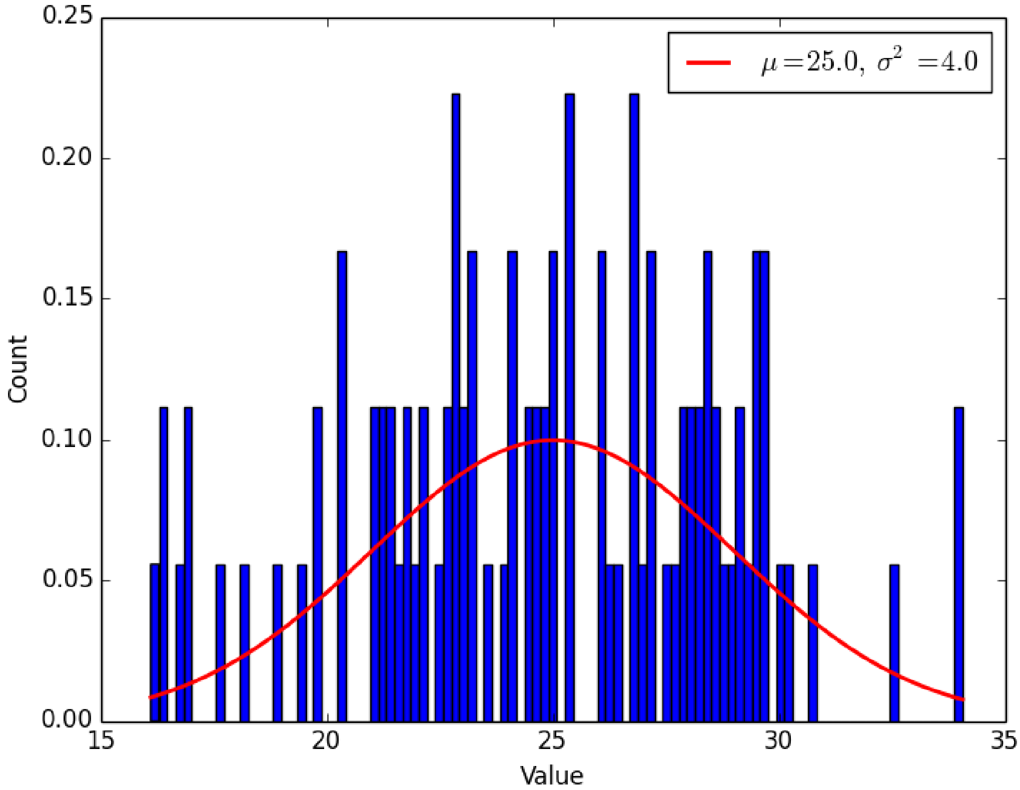


Figure 4.4: Normal distribution applied on the material Tween 80-LQ-(CQ) – polyoxyethylene(20)sorbitan monoeleat.

There are several materials involved that are uncertain and have impact on final production cost. One of the materials are taken here as an example for normal distribution where default value or base price of the material is $0.0213\,€/mL$. Therefore, a range is selected with mean value 25 and standard deviation 4.

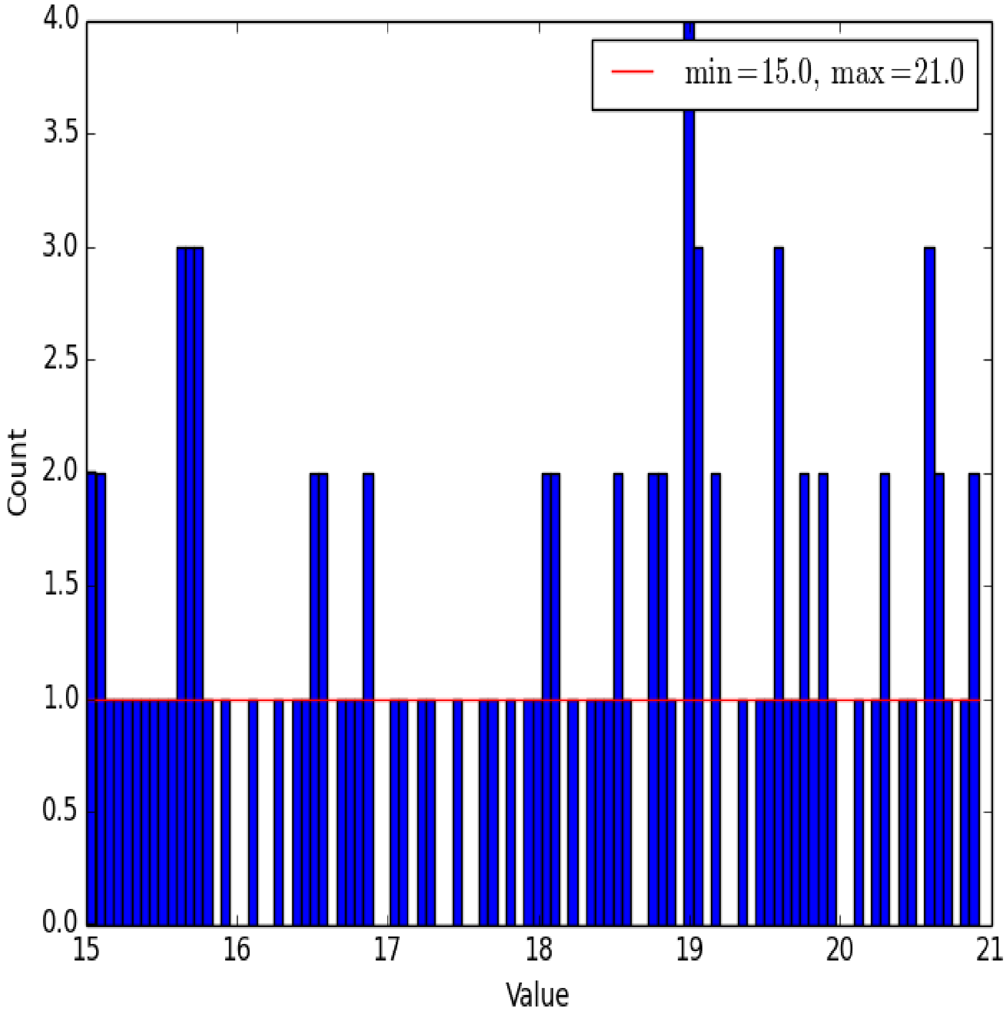


Figure 4.5: Uniform distribution of the equipment Fermenter 30L.

This example of uniform distribution is applied on the equipment Fermenter 30L where L stands for Litter. As equipment usage time has a marginal impact on overall production cost, so it is considered to distribute the usage time to observe the influence closely. Fermenter 30L has the default or base value of 18 hour as usage time. Therefore, a distribution range is being chosen with minimum 15 and maximum 21.

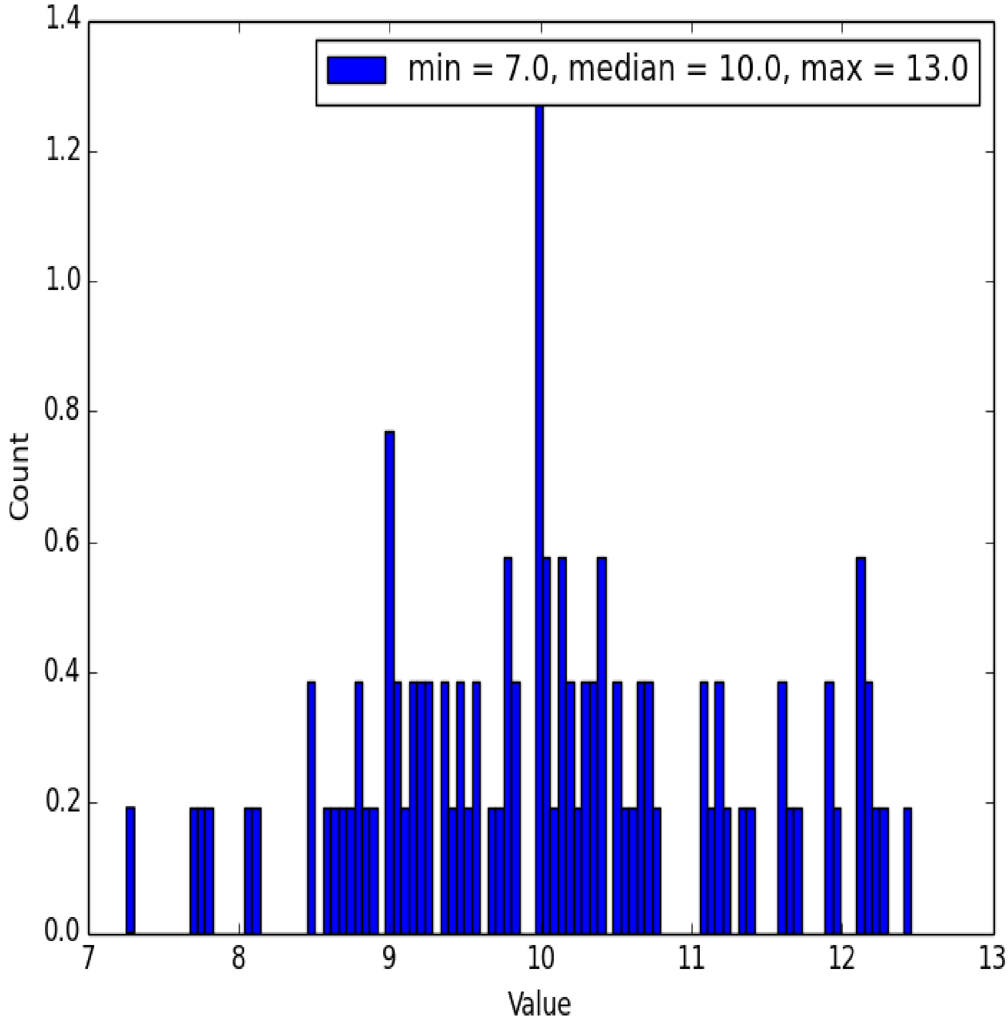


Figure 4.6: Triangular distribution spreaded on the Lyophilizator.

This is the example of triangular distribution. Lyophilizator is another equipment used in the case study where default or base value of it is 10 hour. Hence, a triangular distribution has applied within the range of mean 7, median 10, and maximum 13.

5. CASE STUDY

This chapter emphasizes details about the case study that is been considered and implemented regarding the thesis context. First it talks about the components that the probiotic example comprises of. Later the chapter progresses with the implementation details. The case study example can be divided into different parts which are discussed in following writing.

We would conduct a simulation of the real production scenario to imitate the situation as closely as possible. As we have seen earlier (section 3.3) each scenario consists of several unit processes. Different materials, equipment, and product yields are used in those processes. Hence, our considered case study example i.e. probiotic example consists of following unit processes:

1. Seed fermentation
2. Product fermentation
3. Lyophilization
4. Grinding

Seed fermentation process does not contain any materials and utilities, thus no subsequent cost of those. However, this process uses the equipment Fermenter 30L with the usage time of 18 hour, and final amount of 20L. Uniform distribution is applied here on equipment usage time with minimum value of 15 and maximum value of 21. There are 100 unique values generated for each of those distributions that are going to be used in the simulation. Figure 5.1 shows the measurement setup for the distributions of all the unit process's equipment usage time.

Equipment usage time variation

Unit processes	Equipment	Equipment usage time(h)	Variations			
Seed fermentation	Fermenter 30L	18	Uniform	15	21	
Production fermentation	Fermentor 1500L	18	Uniform	15	21	
Lyophilization	Lyophilizator	10	Triangular	7	10	13
Grinding	Grinder	2	Contant value			

Figure 5.1: List of equipment and their usage time in unit processes.

Process product fermentation uses the equipment Fermenter 1500L. This equipment has 18 hour usage time and the final amount is 20L. In addition, this process contains product yield with an amount of 20 g/L which would be used in the following unit processes. Figure 5.2 depicts the information related to product yields and their relevant distribution. A point to note in the figure that, thus seed fermentation process does not contain any product yield for further processing, that is why the process left uncheck with the empty variation fields. And considering the figure 5.1, likewise the seed fermentation, uniform distribution is considered in product fermentation as well with the same minimum and maximum value.

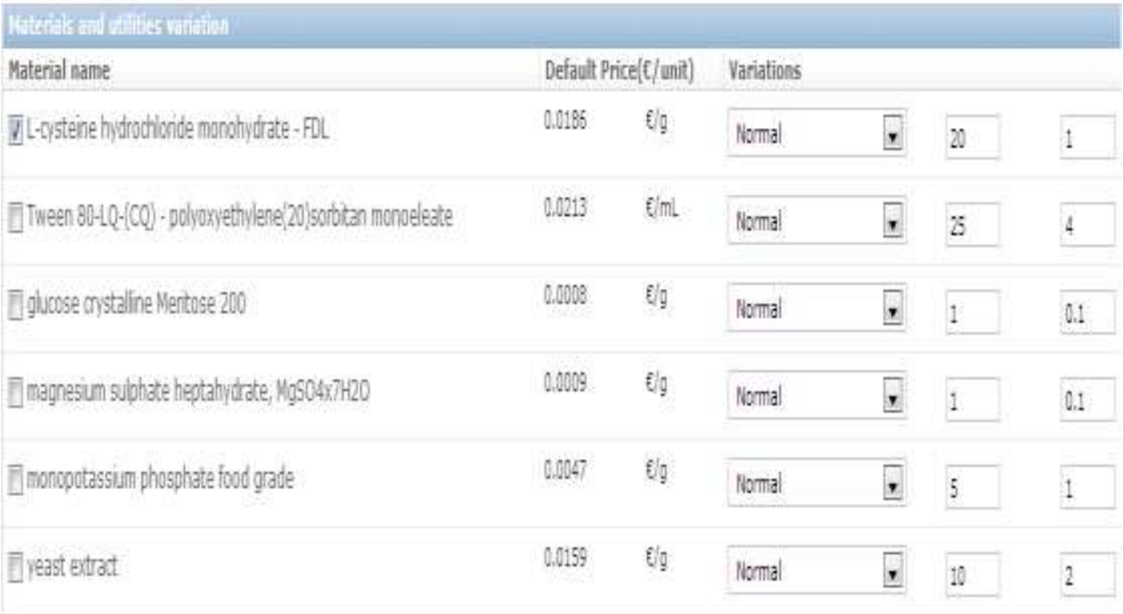
Product yield variation

Unit processes	Product yield		Variations			
Seed fermentation	-	-	Select distribution type			
Production fermentation	20	g/L	Normal	20	3	
Lyophilization	100	%	Uniform	85	100	
Grinding	100	%	Contant value			

Figure 5.2: List of Product Yields and their distribution types

Similarly, lyphilization process would use the equipment lyophilizator with the usage time of 10 hour and final amount of 300.00 kg. The process contains 100% product yield. However, triangular distribution is employed in Lyophilization where values are as follows: minimum – 7, median – 10, maximum – 13 (Figure 5.1).

Finally grinding process uses the equipment grinder which has a usage time of 2 hour, final amount of 300.00 kg and 100 % product yield. There is no distribution attached with this process and variation is set to constant value. Constant value in any variation means that the same value (i.e. default one) would be used throughout the whole simulation while other variables would use their distributed values taking one unique value in each iteration.



Materials and utilities variation

Material name	Default Price(€/unit)		Variations		
L-cysteine hydrochloride monohydrate - FDL	0.0186	€/g	Normal	20	1
Tween 80-LQ-(CQ) - polyoxyethylene(20)sorbitan monoeleate	0.0213	€/mL	Normal	25	4
glucose crystalline Meritose 200	0.0008	€/g	Normal	1	0.1
magnesium sulphate heptahydrate, MgSO4x7H2O	0.0009	€/g	Normal	1	0.1
monopotassium phosphate food grade	0.0047	€/g	Normal	5	1
yeast extract	0.0159	€/g	Normal	10	2

Figure 5.3: List of materials with their pricing and distribution info.

Figure 5.3 above illustrates the information related to material price of the probiotic example. Material cost plays a vital role of uncertainty in the production cost. Therefore, variation of the prices is vital. All the materials belong to the product fermentation process. None other process has any material involved in the simulation. As we can see all the materials have their default price listed with the currency per unit. Here normal distributions are applied to all the materials according to different mean and standard deviation. 100 uniformly distributed unique values are generated for each material. Therefore, when the simulation is being executed, one value from the list is picked each time. This is how the price of all the materials is varied to analyze their uncertainty and sensitivity impact in the final production cost.

Consider the following figure 5.4 to get the functional overview of the simulation process.

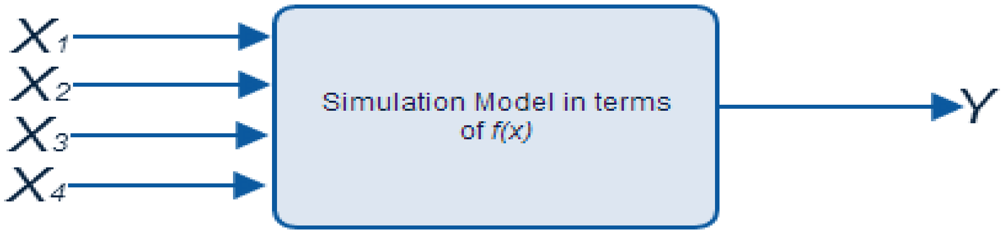


Figure 5.4: Functional overview of how the simulation works.

We have discussed earlier (section 2.4) about the deterministic simulation. Our implemented simulation model is most likely identical to the representation of a mathematical deterministic model i.e. $y = f(x)$. Y denotes the final production cost which is calculated based on the function$f(x)$. And $f(x)$ consists of X_1, X_2, X_3, *and* X_4 where these variables somewhat represent variables such as material cost, equipment usage time, product yield, and fixed cost. Fixed costs for example direct wages, manufacturing cost, administrative overhead, transfer price etc. are added up eventually to the final production cost. X_1, X_2, and X_3, contain 100 distributed values of the variables. When the simulation is executed, it is iterated 100 times and each time one distinct value of X_1, X_2, and X_3, is picked. In each iteration a final production cost is being produced which we represent using tornado diagram and violin plot. Detail analysis of the result would be discussed in the following discussion chapter.

The target of this research is to conduct the simulation model that would be able to help anticipating production cost of a scenario with identification of prominent uncertain variables and their sensitivity. Obtained results from simulation need to be justified to the context in order to accept the sensitivity analysis as credible.

6. DISCUSSION

The results of the case study simulation are presented in this section through histogram, tornado diagram, and violin plot and analytical information attained from is elaborated. One of the objectives of the work was to observe the effects of uncertain variables that helps decision maker to take better decision in the production process. In earlier section (section 2.3) challenges regarding uncertain parameters are mentioned. Hypothesis in the context of decision making tells that identifying those critical variables would help to come up with better decision. Hence, in this discussion section the credibility of the hypothesis and challenges would be assessed. This section can be divided three parts in general. The part contains analysis of the results based on histogram, second part contains result analysis based on tornado diagram and third part talks about the result analysis on the basis of violin plot.

6.1 Histogram Analysis

Figure 6.1 shows the corresponding histogram from the simulation carried out with the distribution information illustrated in figure 5.1, 5.2 and 5.3. The histogram in the figure is the overall histogram defined by the total number of final production cost which is the result of all the uncertain variables varying in their distribution range. The interval of the total cost is 50 labeled horizontally in the figure, and occurrences vertically. The total number of iterations for the simulation model was 100. We can see from the graph, the largest bar is centered at 172246. Similarly quite a lot of smallest bars appeared at different cost, but most of them lie around the center. The histogram does not have any conventional shape such as symmetric, bimodal, skewed to left or skewed to right. This is rather a random one, and the reason is because several distributions such as normal, uniform, and triangular are applied together in the simulation. Therefore we can see there is almost two outlier bars appeared at 173000 and 173048 respectively. These two values cannot be exactly tagged as outliers because we do not have such upper and lower boundaries defined in the simulation. One point to notice in the histogram is that relatively large bars are close to center.

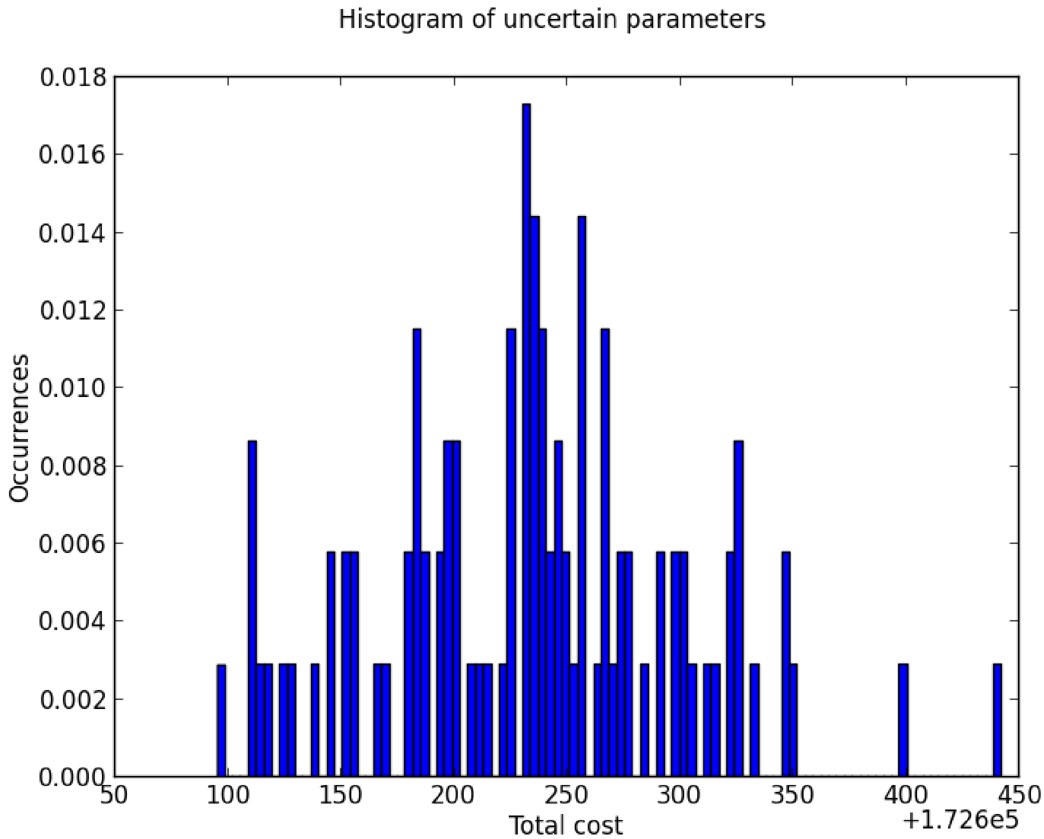


Figure 6.1: Cost of final production and their frequency of occurrences

Though we have the information of final production cost with their frequencies, but the histogram does not really tell us about the impact of the uncertain parameters and their sensitivity. This can be considered as an overall impact scenario but this is not sufficient enough. In this research it is quite important to identify the level of sensitivity of every one of those uncertain parameters. Each and every string needs to be pulled out to extract detail information. Therefore, we would analyze the tornado diagram in the next to section to identify the effects more closely.

6.2 Tornado Diagram Analysis

Tornado diagram provides simple form of sensitivity analysis by varying one parameter at a time in the model and examine the impact that the change appears in the model's output. The analysis could, of course, be repeated on different parameters at different times. We can consider the following figure 6.2.

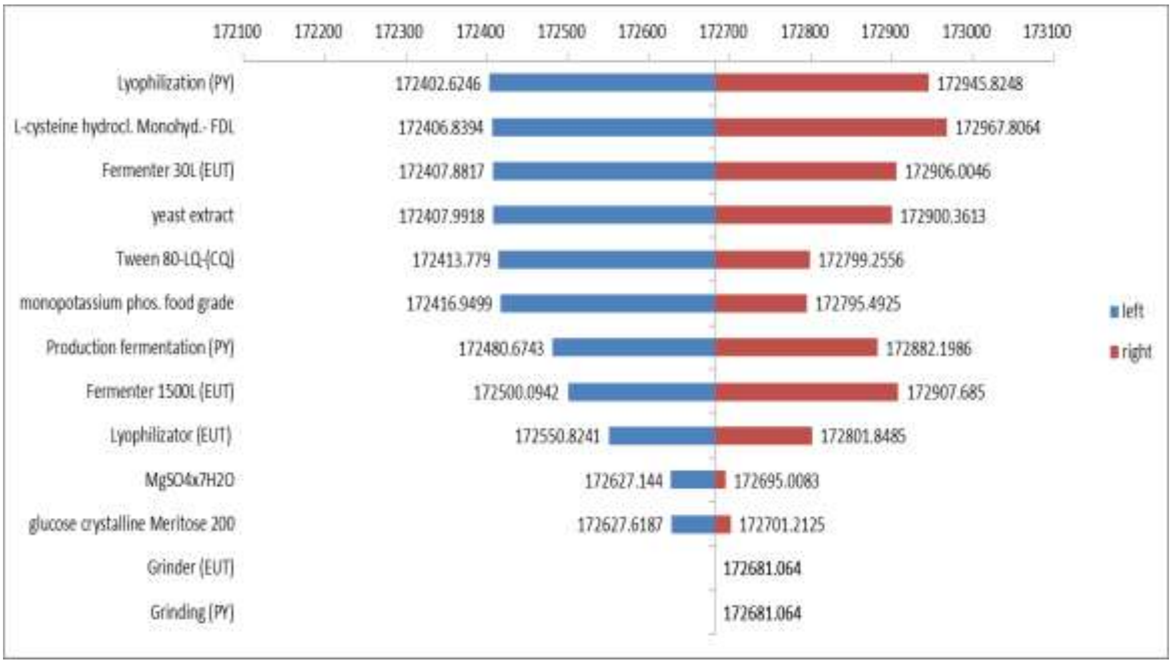


Figure 6.2: Resulting Tornado Diagram of the simulation model.

In the diagram PY stands for product yield and EUT stands for equipment usage time. Base value i.e. final production cost of simulation model using all the parameters with their default values is set to 172681.06 euro. Horizontal bars show the variation occurred in either side due to the distribution of that specific uncertain variable. Maximum and minimum values are entitled in the both side of the bars and base value if there is none. Bars are sorted from largest to smallest on the left side. This diagram means that uncertainty in PY in Lyophilization unit process, L-cysteine hydrochloride monohydrate – FDL, Fermenter 30L, yeast extract, Tween 80-LQ-(CQ) - polyoxyethylene(20)sorbitan monoeleate, and Fermenter 1500L have relatively a big impact on production cost. This indicates that, careful looking should be put at evidence for doing additional research in those specific scopes, focusing on how to keep the upside less driven and downside more optimal. Our main focus is to keep the final production cost reasonable and optimal so that decision making task becomes easier. As we can see material magnesium salphate (i.e. $MgSO_4x7H_2O$), equipment grinder, and PY in Grinding process matters less or not at all than PY in lyophilization in this simulation, so that we can keep them as they are and not to be concerned much. Of course one important point to keep in mind that, above 3 parameter's varying scope were quite narrow. For instance, magnesium salphate was varied by normal distribution with mean 1 and standard deviation 0.1, and grinder and PY in Grinding were set to constant. Obviously they could have noticeable impact if they would be varied in different range. In contrast, magnesium salphate (i.e. $MgSO_4x7H_2O$), lyophilizator equipment, and glucose crystalline Meritose 200 are important in our calculation, though we don't need to focus our attention on further analysis on their current distributions for these factors because the impact of their uncertainty is small and eventually bottom in the order. Hence, this insight brings to be focused on the right factors for further analysis. The TD uniquely identifies those impact factors and uncertainty of variables that are relatively high in both ends. Usually a lot of time and energy is being wasted to identify the high impact uncertain variables if the right way is not approached. The task becomes harder especially when the list of uncertain parameters is comparatively long. Therefore TD comes out with a simple yet powerful solution and provides a graphical representation which becomes handier.

6.3 Violin Plot Analysis

In this section we would analyze the resulting Violin Plot (VP) of the simulation model. Let us look at the following figure (6.3). Three VP is being generated where VP along with MC stands for material cost, VP along with EUT stands for equipment usage time, and VP along with PY stands for product yield. VPs show the median and range (thin vertical line), though it is little harder to look at in this squeezed diagram. Nevertheless we would enlarge to observe individual diagrams more closely. However, diagram shoes that equipment usage time and product yield are little more variant than material cost. Kernel density in VP is discussed earlier (section 4.2), and thus density trace highlights the peaks and valleys in the distribution.

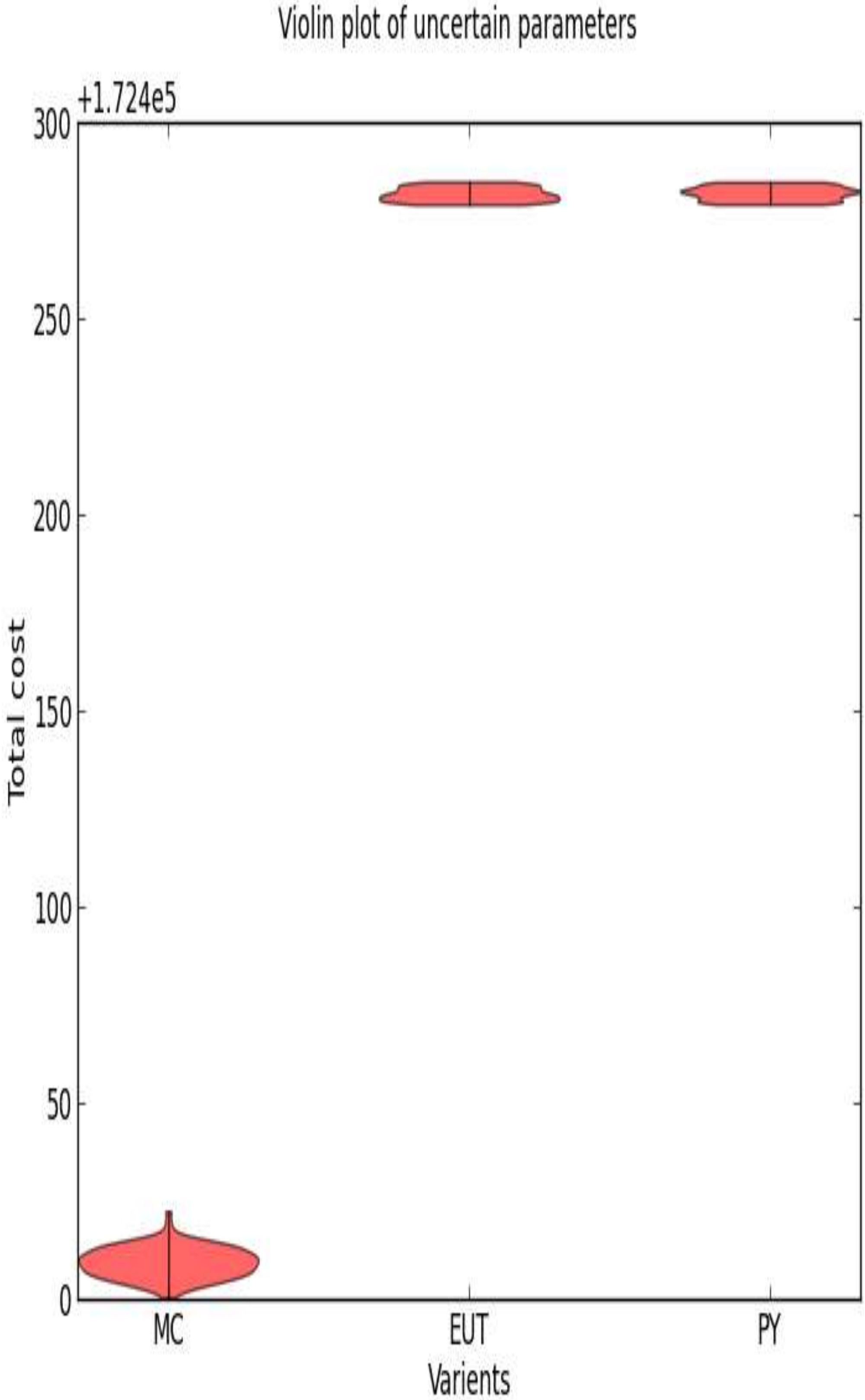

Violin plot of uncertain parameters
+1.724e5
300
250
200
150
100
50
0
Total cost
MC
EUT
PY
Varients

Figure 6.3: Violin Plot produced by the simulation model where MC stands for Material Cost, EUT for Equipment Usage Time, and PY for Product Yield.

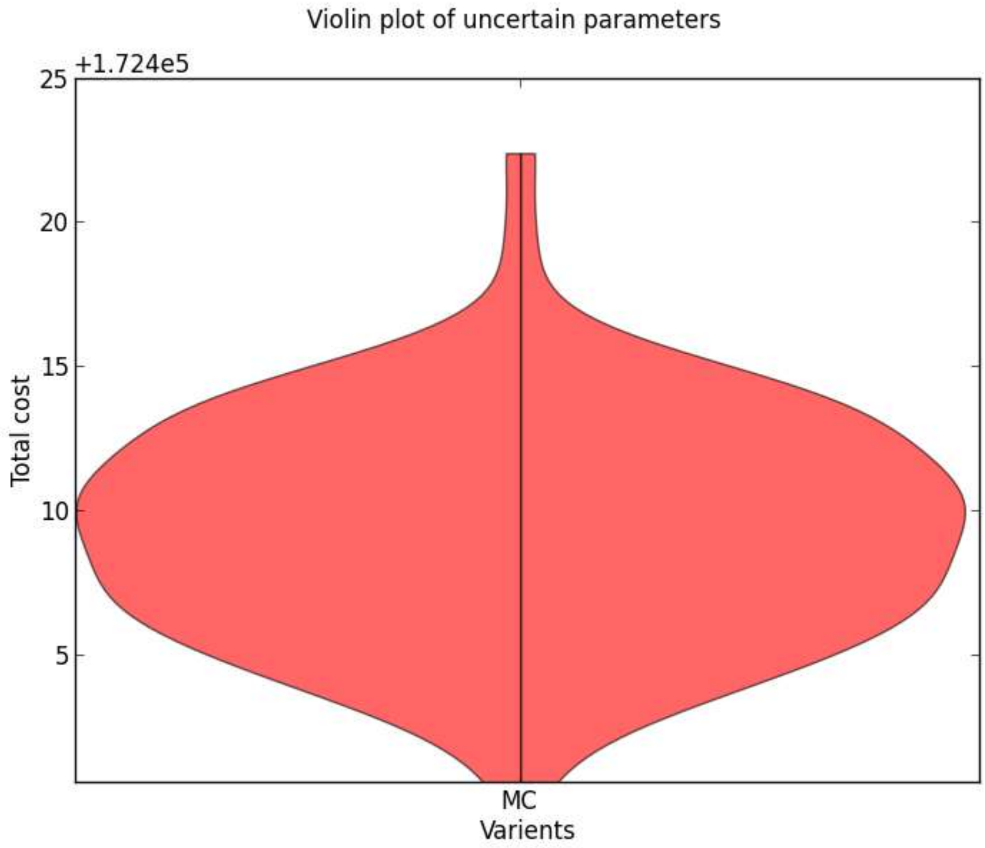


Figure 6.4: Violin Plot of material cost.

Figure 6.4 depicts the VP of material cost. The image has been magnified from the previous figure (i.e. figure 6.3) to be apparent more clearly. Data that is used to generate this VP is shown and elaborated briefly in chapter 5, more specifically in figure 5.3. Even though the range of distributions of each material was different, but type of distribution was set to normal. Therefore, the shape of this VP reveals as the general shape of the normal distribution mentioned here [(Hintze and Nelson 1998)]. One point to notice in figure 6.3 is that other two plots (magnified slice of two plots together is shown in figure 6.5) share almost relatively similar location and scale characteristics, lies on the top of the diagram whereas material cost lies on the bottom of the diagram. Range of the VP of material cost is marked with the thin vertical line and median is at 172410. The trace is added to the violin plot as two symmetric curves on either side of the box plot, making the density and magnitude easy to see. And as expected, distribution shape is being revealed almost accurately by the density trace.

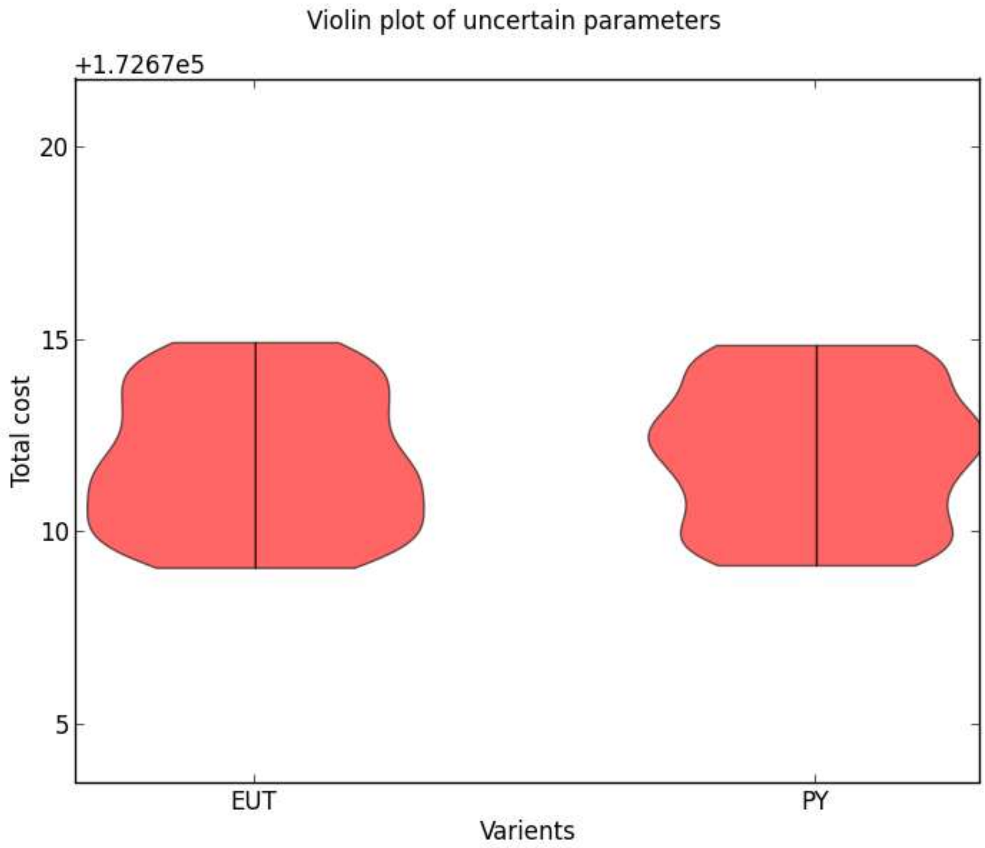


Figure 6.5: Violin Plot of Equipment usage time and Product yield

VP of equipment usage time and product yield is also appeared as expected. Both of the plots have mixed of distribution instead of single type. Equipment usage time has uniform and triangular distributions, and product yield has normal and uniform distributions. Nevertheless, the ability to detect general shapes for distributions of data has produced the expected shapes. These two plots have twin peaks. In equipment usage time, bottom peak has more valley than top peak. That’s why bottom peak is slightly wider than top. On the other plot i.e. product yield, both the peaks are sharper than the plot of equipment usage time. Peaks are opposite compared to the other plot; top peak has slightly wider valley than the bottom peak. Following two figures (6.6

and 6.7) would give closer look of their peaks and valleys as they were sliced and zoomed from figure 6.3.

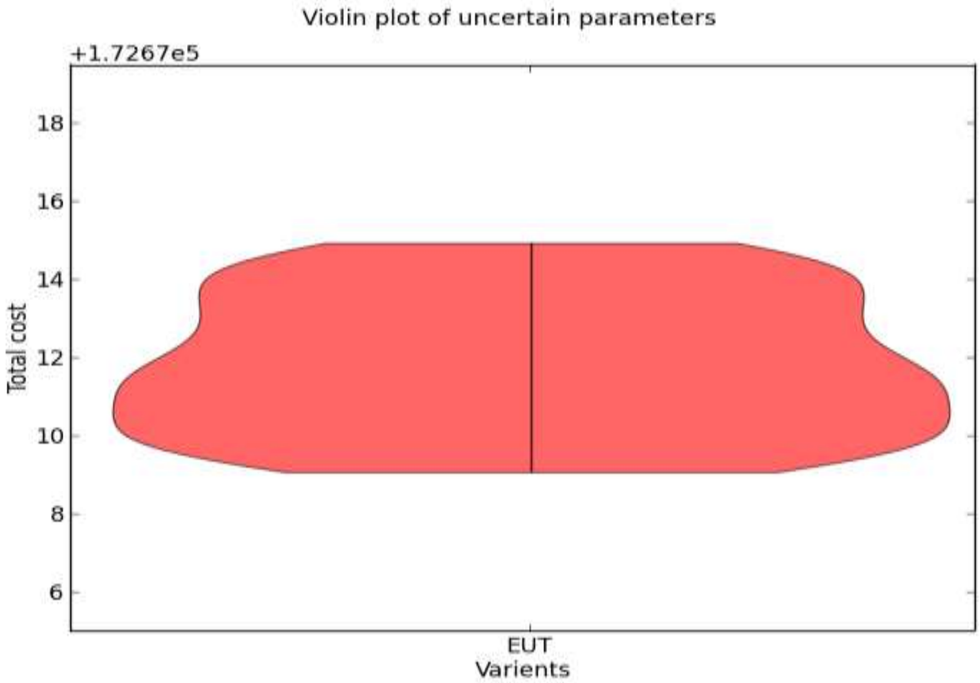


Figure 6.6: Magnified violin plot of Equipment usage time

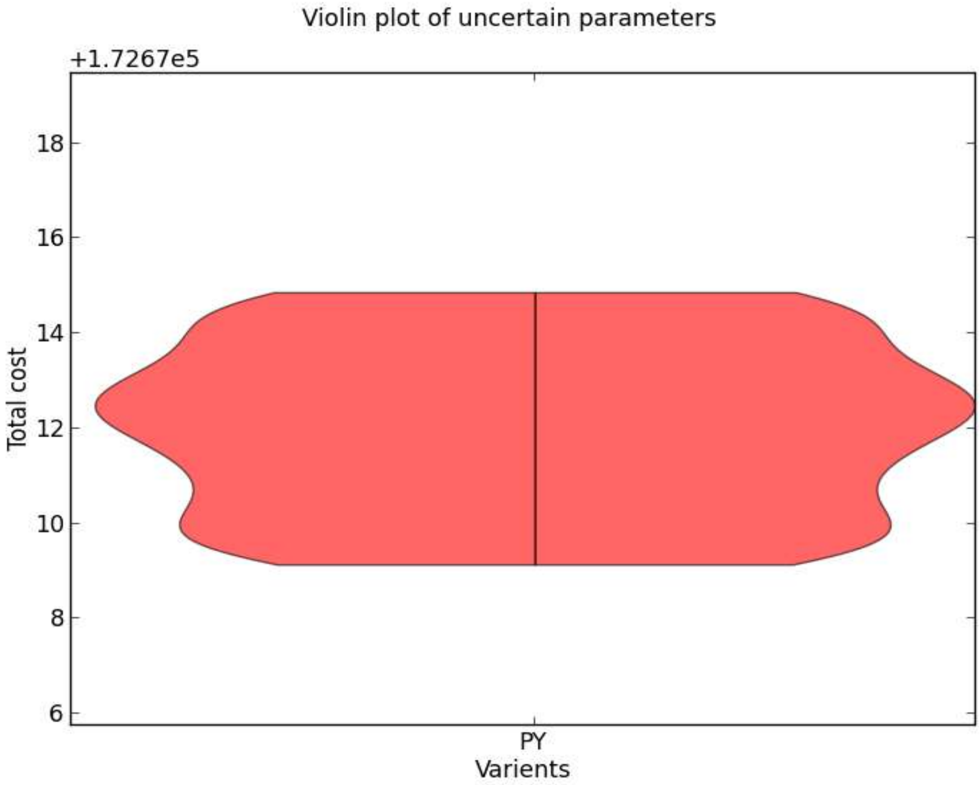


Figure 6.7: Magnified violin plot of Product yield

6.4 Experience

The overall journey was more or less smooth yet few points to be worth mentioning. The tornado diagram was a bit harder to implement comparing to histogram and violin plot. Current mathplotlib library does not have any straight forward way to implement tornado diagram the way we wanted. Built-in library implementation provides labeling in the middle of the diagram with upside and downside of the base value on its right and left respectively. But we required labels (i.e. uncertain variables) on the left and base values (i.e. default price) align in the middle with increase and decrease of each variable from base value on the right and left. Therefore, we had to use a middleware to produce our desired tornado diagram.

7. CONCLUSIONS

The thesis focuses on identifying the impact of uncertain variables and their sensitivity in bioprocessing. A sensitivity analysis tool is been extended from a former tool to achieve the goal. Primary goal is achieved successfully. The Bioptima Planner has assisted to design a simple simulation model where we could mark the uncertain variables and distribute their values in different range. If the tool had to build from scratch instead of extending, it would have made the context more complex to initiate the simulation environment. Though the simulation model was intend to represent the real production scenario to some extent, but there is always some loss in the actual system due to use of machineries which cannot be represent through these pilot models. No system can be found where 100% yield of product would be supplied to the next available process. A minimum level of tolerance would always be present in the system. However it is worthy identify major sources of uncertainty regarding decision making. Thus, the major instigator of sensitivity is variable's uncertainty. The designed model has proved to show us important necessary factors with relevant information and their behaviors. We have illustrated our resulting data using histogram, tornado diagram, and violin plot. Among them, tornado diagram came up with relatively better explanation of information as it depicted variables individual impact. Violin plot has provided group impact. Both of them were worth analyzing as they answer more "what-if" conditions.

Apart from above, there were few limitations in our experiment. As mentioned earlier, a middleware has been used to produce desired form of tornado diagram which caused little delay in displaying all 3 resulting images (i.e. histogram, tornado diagram, and violin plot) and could not be displayed in one go in the final presentation url. That entire data set related to tornado diagram had to pass separately to the middleware in order to produce the diagram. This could be overcome by putting little more time and effort in the implementation of the model. Another limitation was exception handling in the implementation. Current implementation does not handle exceptions, and if the all the necessary inputs are not given then there are several errors rise up such as KeyError, TypeError, ValueError etc. One more point to mention here is that the efficiency of running the simulation. The existing model runs in a reasonable time and takes a little while to display the results. However, code can be optimized to make it little faster though it does not affect the output of the model.

Summarizing all the aspects of the experiment and simulation, the conclusive remark is that the variables those have bigger uncertainty impact and sensitivity is the key scope to be concerned in terms of decision making. The real production scenario should be focused on optimal production cost for highest profitability.

The prime contribution of the thesis is identification of the uncertain variables. The combination of tornado diagram with violin plot is considered as a success of the work. An important observation is that the higher the density trace of the violin plot is, the wider it becomes which gives more valleys to the plots.

8. REFERENCES

Alford, Joseph S. ”Bioprocess control: Advances and challenges.” *Computers and Chemical Engineering* 30 (2006): 12.

Ang, Alfredo H-S, ja Wilson H. Tang. *Probability Concepts in Engineering.* 2003.

Anthony, V., Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* Academic Press, New York, 1983.

Bellman, R., ja K.J. Astrom. ”On Structural Identifiability.” nro 7 (1970): 11.

Bioproduction. ei pvm. http://en.wikipedia.org/wiki/Bioproduction (haettu 11. 10 2012).

Burgess, J.G., R. Osinga, ja R.H. Wijffels. *Marine Bioprocess Engineering.* Elsevier Science, 1999.

Crick, M.J., M.D. Hill, ja D. Charles. ”The Role of Sensitivity Analysis in Assessing Uncertainty.” *Proceedings of an NEA Workshop on Uncertainty Analysis for Performance Assessments of Radioactive Waste Disposal Systems*, 1987.

Critchfield, Gregory C., ja Keith E. Willard. ”Probabilistic Analysis of Decision Trees Using Monte Carlo Simulation.” *Medical Decision Making*, nro 6 (1986): 8.

David, J., Pannel. ”Sensitivity analysis: strategies, methods, concepts, examples.” 1997.

Doran, Pauline M. *Bioprocess Engineering Principles.* 1st. San Diago: Academic Press Inc., 1995.

Doubilet, P, CB Begg, MC Weinstein, P Braun, ja BJ McNeil. ”Probabilistic sensitivity analysis using Monte Carlo simulation. A practical approach.” *Medical Decision Making* 5, nro 2 (1985): 21.

Fery, H. Christoper, ja R. Sumeet Patil. ”Identification and Review of Sensitivity Analysis Methods.” 2002: 26.

Francois, Brun, Wallach Daniel, Makowski David, ja W. Jones James. *Working with dynamic crop models: Evaluation, Analysis, Parameterization, and Application.* Elsevier Science, 2006.

Hamby, D.M. ”A review of techniques for parameter sensitivity analysis.” *publication in Environmental Monitoring and Assessment*, nro 32 (1994).

Hintze, Jerry L., ja Ray D. Nelson. ”Voilin Plots: A Box-Plot Density Trace Synergism.” May 1998: 5.

Idefeldt, J., ja M. Danielson. ”A Note on Tornado Diagrams in Interval Decision Analysis.” *World Congress Engineering* I (2007): 4.

Iman, R.L., ja J.C. Helton. ”The Repeatability of Uncertainty and Sensitivity Analyses for Complex Probabilistic Risk Assessments.” nro 11 (1991): 16.

Jochen, Strube, Grote Florian, Pablo, Josch Jan, ja Ditz Reinhard. ”Process Development and Design of downstream Process.” 2011: 22.

Kalos, Malvin H., ja Paula A. Whitlock. *Monte Carlo Methods.* 2nd. WILLY, 2008.

Khan, Firdos Alam. *Biotechnology Fundamentals.* 1st. CRC Press, 2011.

King, J. M. P., N. J. Titchener-Hooker, ja Y. Zhou. ”Ranking bioprocess variables using global sensitivity analysis: a case study in centrifugation.” *Bioprocess Biosyst Eng*, 2007: 12.

King, J.M.P., N.J. Titchener-Hooker, ja Y. Zhou. ”Identification of key bioprocess variables using global sensitivity analysis.” Springer, 2007.

Monte Carlo mathod - Mathworld. 2103. http://mathworld.wolfram.com/MonteCarloMethod.html.

Monte Carlo SImulation - MATLAB and SIMULINK. 2013. http://www.mathworks.se/discovery/monte-carlo-simulation.html.

Monte Carlo Simulation and Methods Introduction - Goldsim. 2013. http://www.goldsim.com/Web/Introduction/Probabilistic/MonteCarlo/.

Mukhopadhyay, S.N. *Process Biotechnolgoy Fundamentals.* second. New Delhi: Viva books private limited, 2004.

Python. *Python Programming Language - Official Website.* ei pvm. http://www.python.org (haettu 2012).

Rubinstein, Reuven Y., ja Dirk P. Kroese. *Simulation and the Monte Carlo Method.* 2nd. WILEY, 2008.

Saltelli, A., ym. *Global Sensitivity Analysis: The Primer.* 1st. WILEY, 2008.

SHlOYA, SUTEAKI, KAZUYUKI SHlMlZU, ja TOSHIOMI YOSHlDA. ”Knowledge-Based Design and Operation of Bioprocess Systems.” *BIOSCIENCE AND BIOENGINEERING* 87, nro 3 (1999): 6.

Wang, Yong, Ju Chu, Yingping Zhuang, Yonghong Wang, Jianye Xia, ja Siliang Zhang. ”Industrial bioprocess control and optimization in the context of systems biotechnology.” *Biotechnology Advances*, 2009: 7.

What is a deterministic model? ei pvm. http://courses.ttu.edu/isqs5349-westfall/images/5349/deterministic_stochastic.htm (haettu 29. 10 2013).

Zaid, A., H.G. Hughes, E. Porceddu, ja F. Nicholas. ”Glossary of biotechnology and genetic engineering.” 1999: 100.